



# Knowledge Augmentation Services: Virtual Collections & Micro Articles

Kaylin Bugbee

2016 GHRC User Working Group Meeting  
Sept 20-21, 2016

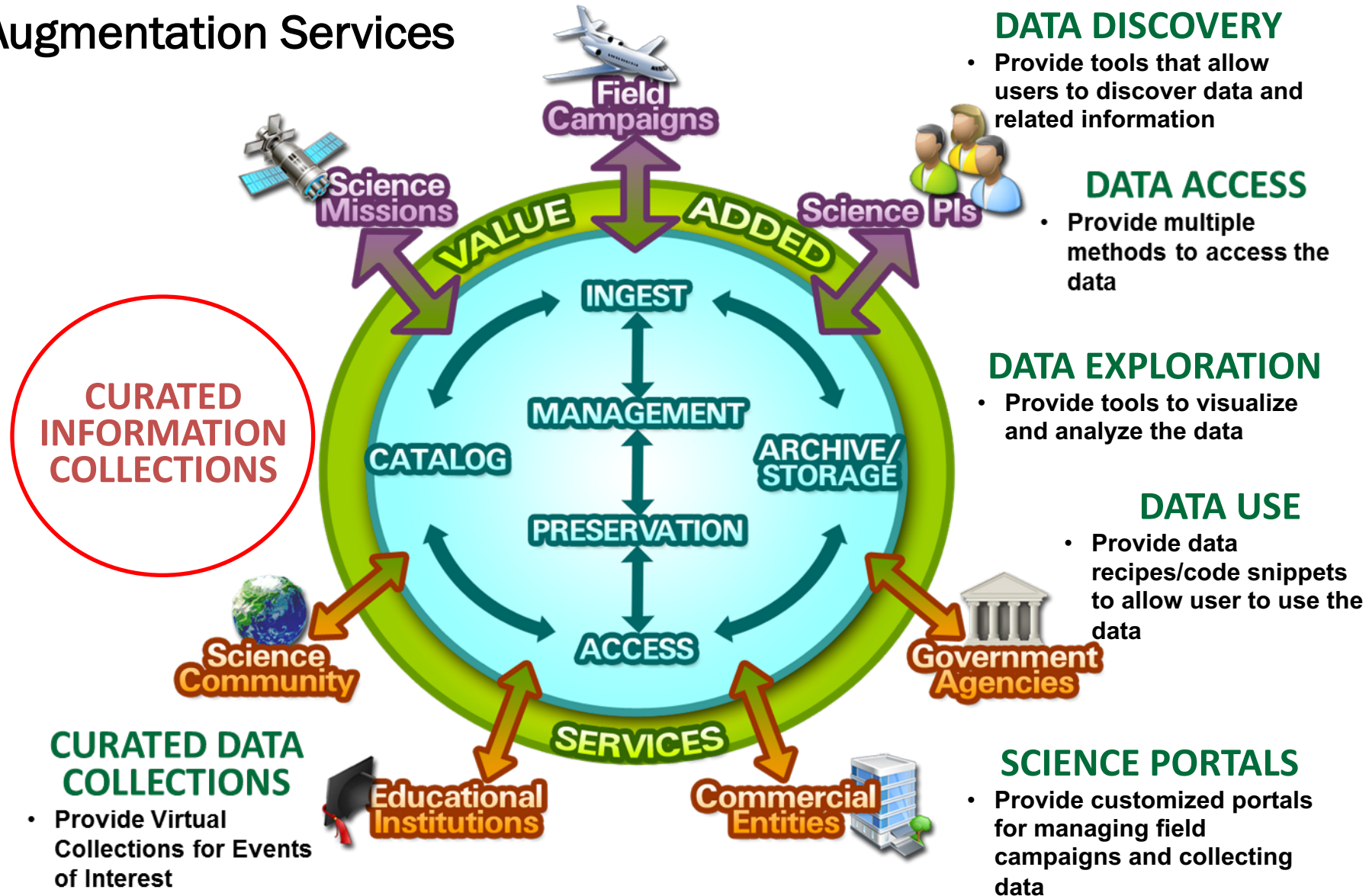


- Broadly speaking, there is a thirst for curated knowledge online
- Examples: Explainer YouTube video channels such as ASAP Science, Minutephysics, TedEd, etc...
- Users appreciate time and effort of putting concepts together by a trusted curator
- In the Earth sciences, “a much larger community of diverse users clamor to access, understand, and use climate data” (Overpeck et al., 2011)
- Initial user community
- Unanticipated users
- The curation and synthesis of Earth science data and related knowledge reduces effort and uncertainties for both the initial user community and unanticipated users and increases certainty of reward

Overpeck, J.T., Meehl, G.A., Bony, S., Easterling, D.R., Climate data Challenges in the 21<sup>st</sup> Century, *Science*, 33, DOI: 10.1126/science.1197869, 700 – 702, 2011.

## Knowledge Augmentation Services

### DATA & SCIENCE EXPERTISE



# Virtual Collections

## What are virtual collections?

- Curation and special collection concept at libraries or museums
- In the digital environment, special collections have expanded and become more flexible
- Known as a virtual collection
- Curated around topic, organization or purpose
- Typically openly available
- Objects can be grouped in various way simultaneously
- The virtual collection is an Earth science data curation service that is growing in use
- Recommendation 17 from UWG requests “Create data bundles for scientists who want to study processes.”
- We define a virtual collection as an end product of a curation activity that searches, selects, and synthesizes diffuse data and information resources around a specific theme/topic or event



## Use Case:

Created in collaboration with Dr. Patrick Gatlin  
(NASA MSFC)

Global Precipitation Measurement (GPM) mission  
Cold-season Precipitation Experiment (GCPEX) :

Synoptic scale snow event that occurred on February  
24, 2012 across southern Ontario

## Resource used to identify event:

Gail Skofronick-Jackson, David Hudak, Walter Petersen, Stephen W. Nesbitt, V. Chandrasekar, Stephen Durden, Kirstin J. Gleicher, Gwo-Jong Huang, Paul Joe, Pavlos Kollias, Kimberly A. Reed, Mathew R. Schwaller, Ronald Stewart, Simone Tanelli, Ali Tokay, James R. Wang, and Mengistu Wolde, 2015: Global Precipitation Measurement Cold Season Precipitation Experiment (GCPEX): For Measurement's Sake, Let It Snow. Bull. Amer. Meteor. Soc., 96, 1719–1741, doi: 10.1175/BAMS-D-13-00262.1.



## Use Case Data Search and Select Process -

1. *Formulate a science question*
  - a. What is the 3-D structure of falling snow and how does its variability affect remotely sensed retrievals?
2. *What are the general keywords related to this science question that can be used to find the dataset required to help answer the question?*
  - a. Remote sensing, snow
3. *What measurements are related to these keywords and what instruments take these measurements?*
  - a. Microwave remote sensing
    - a. Radar: D3R (ground), APR-2 (airborne)
    - b. Radiometer: Dual polarization radiometer (ground)
  - b. Snow particle size
    - a. Cloud microphysics probes (airborne)
  - c. Snow water equivalent
    - a. Nevzorov probe (airborne)
4. *Which Intensive Observing Period (IOP) of the campaign was most of these measurements collected?*
  - a. IOP must consist of a widespread and long duration snowfall event
  - b. IOP must consist of measurements from at least ground and airborne platforms to address spatial variability
5. Data was selected from above instruments

## Synthesis (Implementation):

1. iPython notebook
  - a. Used an iPython notebook to create a virtual collection
  - b. Application environment which allows interactive mode using a web browser
  - c. Provides a distribution URL
  - d. OPeNDAP based aggregation
2. Metadata
  - a. Created test metadata record for this virtual collection in NASA's Common Metadata Repository (CMR)
3. Micro Article
  - a. Short, interesting document that brings together data and key science concepts
  - b. Describes golden case event and relevant data
4. GHRC Publication Workflow
  - a. DOI  
(<http://dx.doi.org/10.5067/GCPEXCS/MULTIPLE/DATA101>)
  - b. Called 'Case Study Collection' in Hydro



**GPM Ground Validation GCPEX Snow Microphysics Case Study**

The GPM Ground Validation GCPEX Snow Microphysics Case Study characterizes the 3-D microphysical evolution and distribution of snow in context of the thermodynamic environment observed during the February 24th, 2012 event of the GPM Cold-season Precipitation Experiment (GCPEX). The GPM Cold-season Precipitation Experiment occurred in Ontario, Canada during the winter season of 2011-2012. GCPEX addressed shortcomings in the GPM snowfall retrieval algorithm by collecting microphysical properties, associated remote sensing observations, and coordinated model simulations of precipitating snow. This case study includes data from the Airborne Second Generation Precipitation Radar (APR-2), Dual-frequency Dual-polarized Doppler Radar (D3R), Dual Polarization Radiometer and the NCAR Cloud Microphysics Particle Probes.

Please include the following citation in your publications:

Gatlin, Patrick N, Manil Maskey and Kaylin Bugbee. 2016. GPM Ground Validation GCPEX Snow Microphysics Case Study. [Indicate subset used]. Dataset available online from the NASA Global Hydrology Resource Center DAAC, Huntsville, Alabama, U.S.A.

DOI: <http://dx.doi.org/10.5067/GCPEXCS/MULTIPLE/DATA101>

[Click here for the citation in RIS format.](#)

[Download Data](#)

In order to increase security for NASA EOSDIS users and to provide a more uniform experience across data offerings, EOSDIS is making a change to the way in which NASA Earth Science data are downloaded. You now need to [create and use an Earthdata Login](#) to download or order GHRC data. Please read our [Earthdata Login recipe](#) page for help and additional information.

**General Characteristics**

Publication date: 2016-06-17

Version: 1

Collections: Case Study Collections, GPM-GV GCPEX Products

Projects: GCPEX

Platforms: NASA DC-8, UND CITATION II

Instruments: APR-2, CIPS, D3R, DPR, HVS, PMS 2D-C PROBE

Terms: ATMOSPHERIC PRESSURE, ATMOSPHERIC TEMPERATURE, ATMOSPHERIC WATER VAPOR, ATMOSPHERIC WINDS, CLOUDS, PRECIPITATION, RADAR, SNOW/ICE

Processing level: 1

Format: NETCDF-4

**Coverage**

Location: CANADA

North boundary: 46°

West boundary: -81°

East boundary: -78°

South boundary: 44°

Temporal resolution: DAY

Start date: 2012-02-24

Stop date: 2012-02-25

Red dots or areas indicate coverage range.



## Lessons Learned:

### 1. Search

- a. To identify a relevant event, re-visited publications, campaign blog, case study documents and other information relevant to GCPEX
- b. Science expert (Patrick) kept documentation of ‘golden cases’ during OLYMPEX campaign as the events took place to streamline the ‘select’ process at a later date when the data is published and readily available to start creating virtual collections

### 2. Select

- a. Metadata Quality
  - a. Temporal
    - i. Temporal information provided at collection level only - missing from granule level
    - ii. Required manual selection of granules and subsetting to identified temporal period using OPeNDAP
  - b. Spatial
    - i. Spatial information provided at collection level only - missing from granule level
    - ii. Required OPeNDAP to subset

## Lessons Learned:

### 3. Synthesize

#### a. Data Formats

- i. To use OPeNDAP, granules need to be provided in standard formats such as HDF or netCDF
- ii. Format translation code was required to convert non-standard granules (ASCII, XLS) to netCDF format
- iii. Note: OPeNDAP-based subsets of the data is only possible if there is a grid type defined for the parameter that we need to subset

#### b. Processing Levels

- i. Granules with different processing levels were served under same dataset
- ii. Parameter-based subset could not be generalized for all the granules for a dataset with same spatial and temporal constraints

#### c. System

- i. IPython notebooks could not be directly hosted within the DAAC due to security concerns. Therefore, only a static (already executed) version of the notebook is being hosted

#### d. Other

- i. Data bundling, Science keyword/vocabulary, Metadata creation in CMR



## Discussion

1. Do you think virtual collections/case study collections will make finding and using data easier?
2. Would you be interested in collaborating with the GHRC on creating new virtual collections?

2016 GHRC User Working Group Meeting  
Sept 20-21, 2016



# Micro Articles

## What are micro articles?

- Micro articles (Tero et al., 2008) were originally defined as academic texts which are longer than a normal abstract but may be one tenth the size of a normal journal article.
  - Focus on gaps in the scientific research process that are generally not documented, published or shared.
- We have defined a micro article as a short, interesting document that brings together data and key science concepts. These documents are curated by both Earth and data scientists.
  - Short, 1 page 'Cheat Sheets' focused on our data holdings
- Micro article types at GHRC include:
  - Physical phenomena related to the data
  - Instrument used to collect the data
  - Event or Case Study (Golden Case)
  - Summary of a cornerstone publication concerning the data
- <https://ghrc.nsstc.nasa.gov/home/micro-articles>

Tero, Heiskanen, Kokkonen Juhana, A Hintikka Kari, Kola Petri, Hintsa Timo, N Pirjo, and Kki. 2008. "Tutkimusparvi: The Open Research Swarm in Finland." *Proceedings of the 12th International Conference on Entertainment and Media in the Ubiquitous Era*. Tampere, Finland: ACM.  
doi:<http://doi.acm.org/10.1145/1457199.1457233>.



## Micro Article – Event

### • Key Information Included:

- Event description
- Description of science behind the event
- Spatial and Temporal coverage
- Access to virtual collection created for the event (DOI)
- Data recipe
- Information on datasets included in the virtual collection including:
  - Dataset access
  - Link to guide
  - Data format information

### • URL:

<https://ghrc.nsstc.nasa.gov/home/micro-articles/snow-microphysics-event-during-gcpex-field-campaign#>

The screenshot shows the GHRC Micro Article page for the "Snow Microphysics Event during GCPEx Field Campaign". The page layout includes a header with the GHRC logo and navigation links. The main content area is divided into several sections:

- Event:** A brief description of the event, stating it occurred on February 24, 2012, across southern Ontario. It mentions that the event consisted of both stratiform and convective bands, and that a significant amount of snowfall was produced. A map of North America highlights the event location.
- Science Description:** A detailed description of the event, explaining that it was part of the GPM Cold Season Precipitation Experiment (GCPEx). It mentions that the event was used to study the relationship between the physical and radiative properties of falling snow. A map of North America highlights the event location.
- Get Data:** A section with a download icon, providing information on how to access the data. It mentions that the data is available in a virtual collection and can be accessed via a DOI.
- Case Study Collection:** A section with a link to the case study collection, providing information on how to access the data.
- Data Recipe:** A section with a link to the data recipe, providing information on how to access the data.
- Dataset Table:** A table listing the datasets included in the virtual collection, with columns for Dataset Name, Guide, and Data Format.
- Relevant Publication(s):** A list of publications related to the event, including a paper by Goff et al. (2013) and a paper by Tashiro et al. (2013).
- Reference Sources:** A list of reference sources, including the GCPEx Case Study Summary Table and the GCPEx Data Recipe.

The footer of the page includes the NASA logo, the date updated (Jul 12th, 2016), the author(s) (Kayla Beggs, Patrick Gatto), the micro article type (Event), and the GHRC logo.

DATASET NAME	GUIDE	DATA FORMAT
GPM Ground Validation Dual Polarization Radiometer (GPR) GCPEx		netCDF
GPM Ground Validation Dual-Frequency Dual-polarized Doppler Radar (DFDR) GCPEx		netCDF
GPM Ground Validation Airborne Second Generation Precipitation Radar (APR-2) GCPEx		HDF
GPM Ground Validation NCAR Cloud Microphysics Particle Probes GCPEx		ASCII

## Micro Article – Publication

- Key Information Included:
  - Publication Citation and DOI
  - Key Findings from the Publication
  - Science Area (GCMD science keyword(s))
  - Datasets Used or Described in the Publication. Includes:
    - Link to dataset landing page
    - Data format
    - Data recipes (if applicable)
    - Science code
- URL:  
<https://ghrc.nsstc.nasa.gov/home/micro-articles/highlights-cecil-et-als-gridded-lightning-climatology-trmm-lis-and-otd-dataset>

GHRC  
Global Hydrology Resource Center

ACCESS DATA MEASUREMENTS FIELD CAMPAIGNS PROJECTS RESOURCES

### Highlights from Cecil et. al.'s 'Gridded lightning climatology from TRMM-LIS and OTD: Dataset description' publication

**Publication Citation**  
 Cecil, Daniel J., Dennis E. Buechler and Richard J. Blakeslee. "Gridded lightning climatology from TRMM-LIS and OTD: Dataset description." *Atmospheric Research* 135-136. (2014): 404–414. doi: 10.1016/j.atmosres.2012.06.028

**Key Findings**

- This publication clearly defines the differences between the 10 different LIS/OTD climatology datasets that are distributed by the Global Hydrology Resource Center (GHRC). Please see the publication for precise details on the methodology used to create each dataset.
- The LIS/OTD 0.5 Degree High Resolution Full Climatology (HRFC) and the LIS/OTD 2.5 Degree Low Resolution Full Climatology (LRFC) are the simplest of the climatology products. Mean annual flash rate is provided as a simple counting experiment – the total number of flashes are divided by the total observation duration.
- Caution is recommended for all data products at high latitudes due to the small sample sizes from OTD.
- The LIS/OTD 2.5 Degree Low Resolution Monthly Time Series (LRMTS) weighs the LIS observations more heavily than the OTD observations since LIS was a higher quality instrument on a more stable platform. The monthly LRMTS are recommended for almost all scientific purposes.

**Science Area**

EARTH SCIENCE > ATMOSPHERE > ATMOSPHERIC ELECTRICITY > LIGHTNING

**Datasets Used**

DATASET NAME	DATA FORMAT	SCIENCE CODE
LIS/OTD 0.5 Degree High Resolution Annual Climatology (HRAC)	HDF	Read SDS Arrays Get Grd Dimensions
LIS/OTD 0.5 Degree High Resolution Full Climatology (HRFC)	HDF	Read SDS Arrays Get Grd Dimensions
LIS/OTD 0.5 Degree High Resolution Monthly (HRMC)	HDF	Read SDS Arrays Get Grd Dimensions
LIS/OTD 2.5 Degree Low Resolution Annual Diurnal Climatology (LRADC)	HDF	Read SDS Arrays Get Grd Dimensions
LIS/OTD 2.5 Degree Low Resolution Annual Climatology (LRAC)	HDF	Read SDS Arrays Get Grd Dimensions
LIS/OTD 2.5 Degree Low Resolution Time Series (LRTS)	HDF	Read SDS Arrays Get Grd Dimensions
LIS/OTD 2.5 Degree Low Resolution Monthly Time Series (LRMTS)	HDF	Read SDS Arrays Get Grd Dimensions
LIS/OTD 2.5 Degree Low Resolution Diurnal Climatology (LRDC)	HDF	Read SDS Arrays Get Grd Dimensions
LIS/OTD 2.5 Degree Low Resolution Full Climatology (LRFC)	HDF	Read SDS Arrays Get Grd Dimensions
LIS/OTD 2.5 Degree Low Resolution Annual Climatology Time Series	HDF	Read SDS Arrays Get Grd Dimensions

## Micro Article – Instrument

### • Key Information Included:

- Instrument description
- Platforms
- Measurements
- Applications
- Spatial resolution
- Swath width
- Accuracy
- Wavelength
- Sampling duration
- Key datasets
- Relevant publications

### • URL:

<https://ghrc.nsstc.nasa.gov/home/micro-articles/earth-observations-lightning-imaging-sensor>

[ACCESS DATA](#)
[MEASUREMENTS](#)
[FIELD CAMPAIGNS](#)
[PROJECTS](#)
[RESOURCES](#)

### Earth Observations: Lightning Imaging Sensor (LIS)

**INSTRUMENT**

Passive Remote Sensing → Spectrometers/Radiometers → Imaging Spectrometers/Radiometers → **Lightning Imaging Sensor (LIS)**

**Description**

The Lightning Imaging Sensor (LIS) detects total lightning (i.e. cloud-to-cloud, cloud-to-ground, and intra-cloud flashes) from a space-based platform.

The LIS is based on digital imaging technology and built around a 128 x 128 charged coupled device (CCD) array that is used to extract only the optical emissions of lightning through Earth's atmosphere for both day and night backgrounds.

**Measurements**

Distribution and variability of total lightning. Amount, rate, and radiant energy of total lightning during both day and night.

**INSTRUMENT PLATFORMS**

Earth Observation Satellites

**Tropical Rainfall Measuring Mission (TRMM)**

Space Stations/Manned Spacecraft

**International Space Station (ISS)**

**Applications**

**Severe Weather**
**Convective precipitation**
**Latent heat**
**Thunderstorms**
**Water & energy cycle**
**Climate**

SPATIAL RESOLUTION	SWATH WIDTH	ACCURACY	WAVELENGTH	SAMPLING DURATION
TRMM LIS: 4 km	Field of View 80 x 80	Detection efficiency ranges from 69% near noon to 88 at night	777.4 nm	approx. 90 s
ISS LIS: 4 - 8 km				

**Key Datasets**

DATASET NAME	GUIDE	SOFTWARE
<a href="#">Lightning Imaging Sensor (LIS) Science Data</a>		<a href="#">Software Library</a> <a href="#">Software Guide</a>
<a href="#">Lightning Imaging Sensor (LIS) Backgrounds</a>		<a href="#">Software Library</a> <a href="#">Software Guide</a>

**RELEVANT PUBLICATIONS**

Boccippio, Dennis J., William J. Koshak, and Richard J. Blakeslee. "Performance assessment of the Optical Transient Detector and Lightning Imaging Sensor. Part I: Predicted Diurnal Variability." *Journal of Atmospheric & Oceanic Technology* 19.2 (2002): 1318.

Mach, Douglas M., Hugh J. Christian, Richard J. Blakeslee, Dennis J. Boccippio, Steven J. Goodman, and William L. Boeck. "Performance assessment of the Optical Transient Detector and Lightning Imaging Sensor." *Journal of Geophysical Research* 112 (2007): D09210, doi:10.1029/2006JD007787.

LIS Algorithm Theoretical Basis Document (ATDB): <http://lightning.nsstc.nasa.gov/bookshelf/pubs/atdb-lis-2000.pdf>

## Published

### **Weather Events**

Snow Microphysics Event during  
GCPEX Field Campaign

### **Publication**

Highlights from Cecil et. al.'s 'Gridded  
lightning climatology from TRMM-LIS  
and OTD: Dataset description'  
publication

### **Earth Observing Instruments**

Lightning Imaging Sensor (LIS)

## Upcoming

### **Phenomena (Type)**

Lightning

Hurricanes

### **Publication**

Highlights from Albrecht et al  
publication

### **Earth Observing Instruments**

Disdrometers

Microwave radiometer



## Discussion

1. Do you think micro articles are beneficial to the GHRC user community?
2. Are there other types of micro articles that you would like to see included?
3. Is there other key information that you would like to see included in a micro article?

2016 GHRC User Working Group Meeting  
Sept 20-21, 2016

